Earable Authentication via Acoustic Toothprint

Zi Wang
Florida State University
Tallahassee, FL, USA
ziwang@cs.fsu.edu

Yili Ren
Florida State University
Tallahassee, FL, USA
ren@cs.fsu.edu

Yingying Chen
Rutgers University
Piscataway, NJ, USA
yingche@scarletmail.rutgers.edu

Jie Yang
Florida State University
Tallahassee, FL, USA
jie.yang@cs.fsu.edu

ABSTRACT

Earables (ear wearable) are rapidly emerging as a new platform to enable a variety of personal applications. The traditional authentication methods thus become less applicable and inconvenient for earables due to their limited input interface. Earables, however, often feature rich around the head sensing capability that can be leveraged to capture new types of biometrics. In this work, we propose ToothSonic that leverages the toothprint-induced sonic effect produced by a user performing teeth gestures for user authentication. In particular, we design several representative teeth gestures that can produce effective sonic waves carrying the information of the toothprint. To reliably capture the acoustic toothprint, it leverages the occlusion effect of the ear canal and the inward-facing microphone of the earables. It then extracts multi-level acoustic features to represent the intrinsic acoustic toothprint for authentication. The key advantages of ToothSonic are that it is suitable for earables and is resistant to various spoofing attacks as the acoustic toothprint is captured via the private teeth-ear channel of the user that is unknown to others. Our preliminary studies with 20 participants show that ToothSonic achieves 97% accuracy with only three teeth gestures.

CCS CONCEPTS

• **Security and privacy → Biometrics**.

KEYWORDS

Biometrics, Toothprint, User Authentication, Earable

ACM Reference Format:
Zi Wang, Yili Ren, Yingying Chen, and Jie Yang. 2021. Earable Authentication via Acoustic Toothprint. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security (CCS '21), November 15–19, 2021, Virtual Event, Republic of Korea.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3460120.3485340

ACM ISBN 978-1-4503-8454-4/21/11.
https://doi.org/10.1145/3460120.3485340

1 INTRODUCTION

Earables are rapidly emerging as a new platform to enable a variety of personal applications due to their rich around the head sensing capability. A recent report shows that earables like Apple's AirPods have much stronger and broader demand than smartwatches and fitness trackers, and are the driving force of the wearable market [7]. Current projections indicate the market for earable devices will reach over 5 billion dollars by 2030, and the earables are becoming smarter and smarter [3]. There also have been increasing research efforts to leverage earables to achieve tasks such as understanding our fitness and sleeps, accessing information, identifying contextual information, monitoring or tracking activities [2].

While earables show considerable promise, they also raise new questions in terms of security. This is because much of the value of the services offered by earables depends on the confidential and personal information they capture, process and transmit. Moreover, earables have also considerable promise as tokens that mediate access to online accounts and diverse devices in smart environments. It is thus critical to developing secure authentication for earables to prevent unauthorized access to security-sensitive data and services.

However, adapting traditional authentication from other wearables or mobiles can be challenging. Quite simply, earables lack a suitable input interface to support rapid and reliable entry of passwords or most of the traditional biometrics. Voice-based authentication is convenient but has been proven vulnerable to voice spoofing attacks [8, 9]. Despite the issue, earables provide novel opportunities to improve or redesign approaches to authentication due to their rich around the head sensing capability. For example, recent work utilizes earable to sense ear canal and its deformation [6] for authentication. However, emitting acoustic sound to probe the ear canal could be intrusive for those who are sensitive to high-frequency sound.

In this work, we propose ToothSonic, a secure earable authentication system that leverages the toothprint-induced sonic effect produced by a user performing teeth gestures for user authentication. In particular, when teeth slide or strike against each other, part of their mechanical energy is released in the form of sonic waves. The harmonics of the friction- and collision- excited sonic wave are dependent on the teeth composition, the dental geometry, and the surface characteristics of each tooth [1]. The key insight is that the sonic waves produce from a teeth gesture carry the information of the toothprint. As every individual has a unique toothprint just like our fingerprint, two users perform the same teeth gesture will result

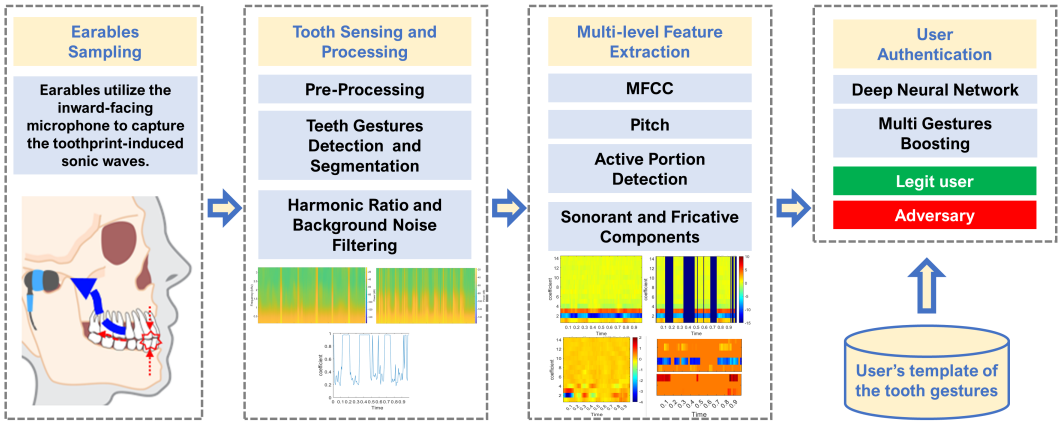


Figure 1: System flow of ToothSonic.

in distinct toothprint-induced sonic waves, which could be sensed by the earables for user authentication. Compared with traditional biometrics, it has several advantages.

Anti-Spoofing. The friction- and collision- excited sonic waves are dependent on the toothprint, which is hidden in the mouth and skull. It is thus more resilient to spoofing attacks compared with traditional biometrics (e.g., fingerprint, face, and voice) that could be exposed to others. In addition, the sonic waves travel through the head tissues and skull channel, which hold the individual uniqueness acting as a hidden and encrypted channel that modulates the sonic waves. ToothSonic is thus resistant to sophisticated adversaries who can acquire the victim's toothprint, for example, via the dentist.

Wide acceptability. ToothSonic provides eye-free and hands-free authentication when hands and eyes are occupied (e.g., carrying objects or driving), whereas most traditional biometric approaches require explicit user operation, such as pressing the fingertip on a reader or posing the face or eyes to the camera. It is also more socially acceptable than voice-based authentication in public places (e.g., offices and libraries) as the sonic waves of teeth gestures are much less perceptible and unobtrusive to others, which also protects user privacy as oppose to audible voice password or paraphrase.

Implicit authentication. ToothSonic can also be exploited as an implicit authentication method when teeth gestures are used as a hands-free computer interface, for example potentially in "Switch Access" services, and for people with motor impairments [4].

In our work, we first design a set of representative teeth gestures based on the factors that impact the toothprint-induced sonic waves. We choose six sliding gestures and four tapping gestures to represent multi-level characteristics of teeth as well as to balance the easy to perform. These gestures can produce effective sonic waves that carry the information of the toothprint. As the sonic waves propagate through the human face and skull to the ear canal, they will be modulated and also significantly attenuated by the teeth-ear channel. To reliably capture the attenuated sonic waves, we choose the inward-facing microphone among various embedded sensors on earables. Utilizing the inward-facing microphone has one key advantage that the earbud and the ear canal form the occlusion effect, which boosts the sonic waves, especially for the low-frequency part that carries effective information of toothprint.

The sensed sonic waves are then going through the pre-processing to remove noise and to segment the data of each gesture. Then, our system extracts fine-grained acoustic features that correspond to the multi-levels of acoustic toothprint, and compares these features against the user enrolled profile to perform authentication. To evaluate ToothSoinc, we conduct experiments with 20 subjects under various scenarios with different number of teeth gestures. Experimental results show that it achieves 97% accuracy with only three teeth gestures.

2 SYSTEM DESIGN

Our system comprises four major components, as shown in Fig. 1. The system uses energy-based event detection to locate the gestures. Next, our system segments the recorded signals into a sequence of gestures by utilizing the Munich Automatic Segmentation system [5]. To enhance the SNR, we apply the harmonic ratio to filter our background noise when no gestures are performing.

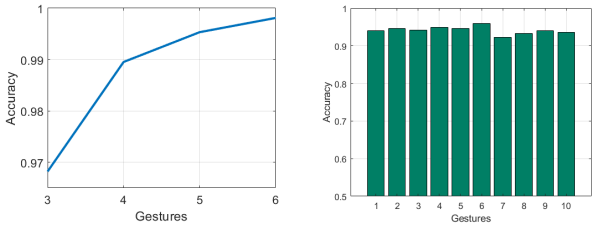
For features extraction, we extract MFCC features (with delta and delta-delta), Pitch, and log spectrum. Also, we locate the active portion where the sonic waves have significant changes. We then extract both sonorant and fricative components for a detailed analysis. Lastly, the user will be authenticated based on the extracted features. Our system adopts a deep neural network to make an authentication decision. Our system can make a decision based on a single gesture. And by leveraging a sequence of teeth gestures, our system could further enhance the authentication accuracy.

As shown in Fig. 2, we design 10 teeth gestures including 6 sliding gestures and 4 tapping gestures. The sliding gestures contain occlusion sliding, molar sliding, canine sliding, incisor sliding front/back, incisor sliding up/down, and incisor sliding left/right. And the tapping gestures are occlusion tapping, molar tapping, canine tapping, and incisor tapping. These gestures cover the major factors that affect the sonic waves of the toothprint when performing gestures.

The first 6 rows mark with blue in Fig. 2 reveal details of 6 different sliding gestures. Sliding to a different direction will reflect the dental mobility toward that specific direction. In addition, all sliding gestures contain information about enamel rod patterns. They could also show dental spacing because teeth generate sonorant sound due to the gaps between each tooth. Meanwhile, the last 4 rows mark with red in Fig. 2 show details of 4 tapping gestures

Biometric Factors / Tooth Gestures	Organ-level				Macro-level							Micro-level			
	Dental Mobility F/B	Dental Mobility U/D	Dental Mobility L/R	Propagation Channel	Dental arch shape	Depth of spee	Occlusion classes	Dental spacing	Incisor shape and size	Canine shape and size	Molar shape and size	Cusp	Enamel thickness	Enamel rod patterns	Tooth root
Occlusion Sliding			√	√	√	√	√	√	√	√	√	√	√	√	√
Molar Sliding	√			√		√		√			√	√	√	√	√
Canine Sliding	√			√				√	√	√			√	√	√
Incisor Sliding F/B	√			√			√	√	√	√			√	√	√
Incisor Sliding U/D		√		√			√	√	√	√			√	√	√
Incisor Sliding L/R			√	√			√	√	√	√			√	√	√
Occlusion Tapping		√		√	√	√	√		√	√	√	√	√		√
Molar Tapping		√		√		√					√	√	√		√
Canine Tapping		√		√		√			√	√	√		√		√
Incisor Tapping		√		√			√		√	√			√		√

Figure 2: Tooth gesture and related biometrics.



(a) Accuracy with multiple gestures **(b) Accuracy for different gesture**

Figure 3: Authentication accuracy.

across different portions of teeth. The sonic waves generated by tapping gestures reveal the mobility on the axis of Z, i.e., up and down. Meanwhile, these sonic waves also disclose the enamel thickness since the enamel dominates the contact area when the tapping happens.

3 PERFORMANCE EVALUATION

We recruit 20 participants for the experiments including 9 females and 11 males with an age range from 22 to 36. For the enrollment, each participant is asked to sit in a living room environment and wear the prototype at her/his habitual position. Next, they are required to repeat each teeth gesture five times. The extracted features of these gestures are used to build the profile for each participant. After enrollment, each participant is asked to repeat each gesture at least 40 times. Each gesture repeat act as one authentication attempt in our experiments.

Fig. 3 (a) shows the accuracy that leverages multiple gestures. In sum, we could see that ToothSonic achieves high accuracy over 99% by combining a few gestures. In particular, our system could achieve authentication accuracy of 99.81%, 99.53%, 98.95%, 96.82% by with 6, 5, 4, 3 gestures, respectively.

Fig. 3 (b) shows the accuracy across 10 different gestures when using only one gesture for authentication. No.1 to No.6 stand for the six different sliding gestures and the left 4 gestures are tapping gestures. We observed that the performance of the sliding gesture is better than the tapping gestures. This is because sliding gestures have a longer duration and contain more tooth participants with different dimensions of information. Therefore, sliding gestures contain more features than tapping gestures, and thus could provide more accurate authentication. We could also find that the accuracy of canine gestures is the lowest. This is due to canine gestures only involve one side canine with less information, and such gestures are harder for users to perform in our experiments.

4 CONCLUSIONS

In this work, we propose ToothSonic, a secure earable authentication system that leverages the toothprint-induced sonic effect produced by teeth gestures for user authentication. ToothSonic has several advantages over traditional biometric authentication including anti-spoofing, wide acceptability, and conditionally implicit authentication. We investigate representative teeth gestures that produce effective sonic waves carrying the information of the toothprint. Multi-level acoustic features are also extracted to represent intrinsic toothprint information. Our preliminary results demonstrate the effectiveness of the ToothSonic in authenticating earable users.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their insightful feedback. This work was partially supported by the NSF Grants CNS-2131143; CNS-1910519; and DGE-1565215.

REFERENCES

[1] Jean-François Augoyard. 2006. *Sonic experience: a guide to everyday sounds.* McGill-Queen's Press-MQUP.
[2] Romit Roy Choudhury. 2021. Earable computing: A new area to think about. In *Proceedings of the 22nd International Workshop on Mobile Computing Systems and Applications.* 147–153.
[3] Idtechex. 2020. . https://www.idtechex.com/en/research-article/what-does-the-future-hold-for-the-hearables-market/22130.
[4] Jay Prakash, Zhijian Yang, Yu-Lin Wei, Haitham Hassanieh, and Romit Roy Choudhury. 2020. EarSense: earphones as a teeth activity sensor. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking.* 1–13.
[5] Florian Schiel, Christoph Draxler, and Jonathan Harrington. 2011. Phonemic segmentation and labelling using the MAUS technique. (2011).
[6] Zi Wang, Sheng Tan, Linghan Zhang, Yili Ren, Zhi Wang, and Jie Yang. 2021. EarDynamic: An Ear Canal Deformation Based Continuous User Authentication Using In-Ear Wearables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (2021), 1–27.
[7] ZDNet. 2020. . https://www.zdnet.com/article/apple-airpods-and-other-smart-hearables-are-ruling-the-wearable-tech-market.
[8] Linghan Zhang, Sheng Tan, Zi Wang, Yili Ren, Zhi Wang, and Jie Yang. 2020. VibLive: A Continuous Liveness Detection for Secure Voice User Interface in IoT Environment. In *Annual Computer Security Applications Conference.* 884–896.
[9] Linghan Zhang, Sheng Tan, Jie Yang, and Yingying Chen. 2016. Voicelive: A phoneme localization based liveness detection for voice authentication on smartphones. In *Proceedings of the 2016 ACM SIGSAC Conference on CCS.* 1080–1091.